International Journal of Engineering Technology and Management Sciences[IJETMS]

Website: ijetms.in Issue:5, Volume No.4, September-2020 DOI: DOI: 10.46647/ijetms.2020.v04i05.014

Toxic Sentiment Identification Using R Programming

G.Divya Bharathi[1], Dr.A.Jagan[2], V.Pradeep Kumar[3]

[1]*M.Tech., B.V.Raju Institute of Technology, Narsapur, Telangana, India.*
[2]*Professor, B.V.Raju Institute of Technology, Narsapur, Telangana, India*
[3]*Assistant Professor,B.V.Raju Institute of Technology, Narsapur, Telangana, India*

[1]divyarevana@gmail.com
[2]jagan.amgoth@bvrit.ac.in
[3]pradeepkumar.v@bvrit.ac.in

Abstract*— Text messaging has become a universal staple. WhatsApp is regularly becoming a news delivery channel as users rely on its broadcast messages to share both local and international news. Today we are not utilizing and operating it, but it is operating us which can confirm to be very unsafe for us. Most of the fake news spread rapidly by WhatsApp. So, there is requirement to examine WhatsApp chat by user's sentiment or opinion. WhatsApp is such an application which is used widely for transferring media, text, files as well as audio calling. WhatsApp is progressively becoming a turning point in numerous sectors like healthcare, education and business. So, there is requirement to inspect WhatsApp chat by user's sentiment or opinion. The advent of the internet had played a huge role in expanding the usage of text messaging to instant messaging on mobile devices. WhatsApp chat sentiment analysis to increase improved insights regarding their employees and strive to stay away from unanticipated conflicts due to various redundancies and insufficiency of business processes. Sentiment analysis is most popular branches of textual analytics which with the aid of information and natural language processing observe and categorize the unorganized written data into different sentiments. It is as well as acknowledged as opinion mining. Most of the false news increase rapidly by WhatsApp. Therefore, there is call for to observe and examine WhatsApp chat to find user's sentiment or opinion. Firstly, chat from WhatsApp is selected and exported to a system which is an easy task and can be done either by phone or WhatsApp for the computer system. Following this, the processes are fairly simple and have been explained with all the coding details needed to analyze the texts. In this project, chat of WhatsApp has been used as database by using R, sentiments and emotions are being analyzed.

***Keywords*—Data Analysis, Identification, R Programming, Sentiments, Whatsapp.**

I. INTRODUCTION

Text messaging has become a universal staple. In today's world, the largest part of accepted chat application for speedy communication is WhatsApp. Every smart phone user utilizes this kind of mobile applications for textual communication. The world is moving towards a fully digitalized economy at an incredible pace and as a result, a generous amount of data is being produced by the internet, social media, smart phones, tech equipment and many others.

Description

It is at no cost and extremely fast communication application of mobile, but in present days public have become addicted to this application and the negative feature of it is as well clear that few persons have started utilizing this for annoying people today. Sentiment analysis is one of the majority accepted branches of textual analytics by the aid of information and natural language processing observes and categorizes the unorganized textual information into different sentiments. It is as well acknowledged as opinion mining as the main focus is on the opinion and approach of the persons by analyzing their textual messages. A quantity of false news increase rapidly by WhatsApp too. So, there is requirement to examine WhatsApp chat by user's sentiment or opinion. Sentiment analysis of WhatsApp chat is done to increase improved insights towards their employees and attempt to stay away from unanticipated conflicts due to various redundancies and insufficiency of business processes. WhatsApp is steadily fetching a news delivery channel as users rely on its broadcast messages to share both local and international news. In this paper, the chat in the group of WhatsApp is taken as database to analyze sentiments and emotions using R Studio.

Motivation

In this project, the analysis of sentiments or emotions of a person are calculated at that particular context. The chat in the documentation which is taken from the WhatsApp is extracted and stemmed. The extracted document is analyzed in according to the occurrence of the words and the sentiments that are present in extracted data. The wordcloud is generated using the frequency of the words and the analysis of the sentiment. This analysis is represented with the assistance of bar graph.

Problem Identification

WhatsApp is such an application which is used widely for transferring media, text, files as well as audio calling. WhatsApp is steadily fetching to be a game changer in various sectors like healthcare, education and commercial applications. So, there is requirement to examine WhatsApp chat by user's sentiment or opinion.

International Journal of Engineering Technology and Management Sciences[IJETMS]

Website: ijetms.in Issue:5, Volume No.4, September-2020 DOI: DOI: 10.46647/ijetms.2020.v04i05.014

Sentiment analysis is solitary of the mainly popular branches of textual analytics which with the aid of information and natural language processing analyze and categorize the unorganized text data to different sentiments. It is as well known as opinion mining.

.

Fig.1.1 Sentiment Analysis

For Example, as per the performed analysis and visualizationB V Raju Institute of Technology, Narsapur, Medak Dist., Telangana the most active day of the week is "Monday" with sum amount of messages sent as well as received are 520.So maximum participation of the senders on WhatsApp group chat takes place on Monday.

To find which age group participants are more active on WhatsApp group and number of messages send by each age group participants per month, day, hour. For Example, the age group of the participants in the Dataset ranges from 14 to 68. So, the youngest person of the group is of age 14 and the elder person has an age of 68.

The sum of the number of messages sent by each age group participants per hour/per day/per month.

II LITERATURE SURVEY

In today's internet era, the fastest communication through mobile application is WhatsApp, in which each smart mobile phone user is sharing their thought, sentiment, and opinion with each other. In this paper, it is investigated group-based communication in WhatsApp.

It remains for future work to analyze the benefits and to study the applicability of each approach in order to adapt the current Internet technologies [1].

Hybrid approach had till now exhibited positive sentiment of the concerned performance. Even though they have been deployed using unigrams and diagrams, their performance is worse on trigrams. This definitely leaves researchers to explore the terrain. Sentimental analysis is a method by which the statement can be analyzed from large text and this method is also known as opinion mining [2].

Other key distinguishing features are the diversity of data sources that one frequently encounters in data mining projects, in addition to the diversity of data types (text, sound, video, etc.). These kinds of all issues turn data mining to an extremely interdisciplinary subject involving not only data analysts but also the people working with various databases, visualization of data on high dimensions, etc [3].

This has altered the way in which public converse and control political, social and economic manners of additional persons in the Web 2.0. Undeniably the Web 2.0 permits everybody who ever has a voice, capable to increase human partnership abilities on international scale, facilitating individuals to contribute to views by the procedure of read-write Web and user's produced contents [4].

The first fault is that because persons can freely share their content of their own, the superiority of their opinions is not guaranteed. Example, as an alternative of allocation of topics associated opinions, spam on forums are being posted by online spammers. Few spams were meaningless at every one whereas others have unrelated views also known as false opinions [5].

Sentiment analysis is the put into practice of relating to natural language processing and message analysis techniques to recognize and take out subjective in order from text. This paper works on a study on the sentiment analysis confront appropriate to their move towards and procedure. It also separates the tests into two kinds to effortlessness to arrangement with them and spotlight on the amount of precise meaning. This research discusses these sentiment challenges, the factors affecting them, and their importance [6].

It is too identified that human analysis of text information is subject to considerable biases, e.g., persons frequently disburse superior attention to opinions that are steady with their have first choice. Supervised machine learning techniques have shown better performance than unsupervised machine, earning techniques. However, the unsupervised techniques are significant too since supervised methods insist huge amounts of labeled preparation information that are very elite while attainment of unlabelled data is easy [7].

In a lot of applications, it is significant to regard as the circumstances of the text and the user choices. This is the reason for the require to make more research on background based SA. By means of TL methods, we can utilize connected data to the field in query as a preparation data. By means of NLP tools to strengthen the SA process has paying attention researchers newly and still requirements some improvements [8].

The linguistic and semantic move toward implemented in this scheme facilitates the research, the investigation, the categorization of huge volumes of heterogeneous credentials, assisting documental analysts to cut from side to side the in sequence labyrinth, analysts to obtain explanation of difficulty of community views, conveying repeatedly a sentiment division, quickly right to use all the latent texts of notice [9].

International Journal of Engineering Technology and Management Sciences[IJETMS]

Website: ijetms.in Issue:5, Volume No.4, September-2020 DOI: DOI: 10.46647/ijetms.2020.v04i05.014

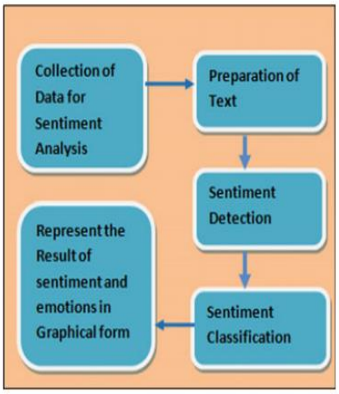

Fig.2.1 Process of Sentiment Analysis

From this survey, it can be concluded that supervised techniques provide better accuracy compared to dictionary based approach. In future, various opinion summarization algorithms should be applied to generate summary of all reviews provided by users [10].

The data used in this study is online product reviews collected from the sample website that we have created. Words of adverbs and adjectives are able to express contradictory sentiment with the aid of pessimistic prefixes. Negation phrase identification algorithm is used to find such words. The performance is evaluated through evaluation measures [11].

It is as well initiated that various kinds of features and categorization algorithms are joined in a well-organized way to prevail over person drawbacks and advantage from every other's virtues, and lastly improve the sentiment classification presentation [12].

III DESIGN METHODOLOGY

Process of cleaning, transforming, inspecting and modeling information with the objective of uncovering useful information, indicating conclusions, and thus supporting decision-making is Data Analysis. Data Analysis process includes Data Collection, Data Transformation, Data Loading, Exploratory Data Analysis, Data Visualization and Report. Data is collected and examined for testing assumptions or answering the queries.

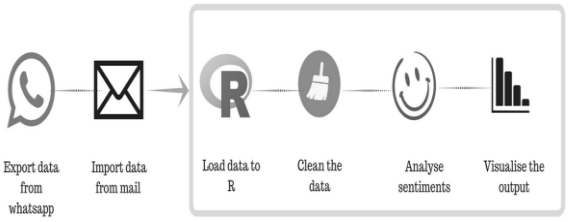

Fig.3.1 Process of Data Analysis

To check whether males are addicted to the groups in the WhatsApp or the females. The frequency of usage of WhatsApp group by males and females. For example, according to the analysis, Females are likely to be additional addicted to the WhatsApp group as compared to Males. It is found that the sum numbers of messages sent by the Females are 2811 and by Males are 2752. This clearly concludes that Females are more involved in the Group.

If used positively then it's a boon for the users and if addicted after that a ban and therefore this research paper classified the level of addiction of users to the WhatsApp chat so as to decrease the time spent on it and to explore the chat whenever necessary.

IV IMPLEMENTATION

Exporting and importing the data from WhatsApp to email

In this case the data been extracted is historic data that holds the key to understand data over time. The aim of this project is to predict whether a particular individual is said to be addicted to WhatsApp group or not and this is done using the R statistics software program.

Collection of data is the primary stage of the model which includes formation of idea, defining the project aim, setting up the machine and lastly understanding the data. Collection of data is necessary to make sure the truthfulness of research.

Loading the data into R

Once R software is installed we can decide to effort with an integrated development environment (IDE) RStudio. It is the mainly popular IDE for R software and aids debugging, management of workspace, plotting and many more. The RStudio window is classified into four panes namely source pane, console pane, workspace pane and plots pane.

Data cleaning

Data cleaning is also known as Data Scrubbing in which imprecise records commencing an exacting dataset are corrected and removed. The reason of cleansing the data is to find out wrong, unrelated or inadequate elements of the information to either change or delete it to make sure that the given data set is precise and reliable with other sets in the system. In case of data cleansing, avoid blank spaces, values or fields; else every word will be treated as a separate variable, which in errors that are related to the number of elements in each line given data set. Avoid special symbols such as @, #, $, ^, ***, ,(,) ,- , ?,,,< ,> , / , | , \ , [,] ,{ , and }. Characters in uppercase are also converted into lowercase characters.

The libraries that are used for data analysis and data cleaning are:

- Library (readtext): **readtext** is a one-function package that does exactly what it says on the tin. It reads files containing text, along with any associated document-level metadata. **readtext** accepts file masks, so that you can specify a pattern to load multiple texts, and these texts can even be of multiple types.
- library (lubridate): Date-time data can be frustrating to work with in R. R commands for date-times are

Website: ijetms.in Issue:5, Volume No.4, September-2020 DOI: DOI: 10.46647/ijetms.2020.v04i05.014

generally unintuitive and change depending on the type of date-time object being used. Moreover, the methods we use with date-times must be robust to time zones, leap days, daylight savings times, and other time related quirks, and R lacks these capabilities in some situations.

- library(tm): A framework for text mining applications within R.
- library (tmap): Thematic maps are geographical maps in which spatial data distributions are visualized. This package offers a flexible, layer-based, and easy to use approach to create thematic maps, such as choropleths and bubble maps.
- library(dplyr): dplyr is a grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges
- library (SnowballC): An R interface to the C 'libstemmer' library which equips Porter's word stemming algorithm for collapsing words to a general root to help comparison of vocabulary. Currently supported languages are English, Danish, Dutch, French, Hungarian, Russian, Romanian, Swedish, Finnish, German, Italian, Norwegian, Portuguese and Turkish.
- library (stringr): Strings are not exciting, high-status components of R, but they do provide a major role in numerous data cleaning and preparation works.

Analysis of Sentiments

Sentiment Analysis is the procedure of determining whether a portion of text is negative, positive or neutral**.** A sentiment analysis system for textual analysis unites natural language processing. The analysis of sentiments is represented to form a required shape of wordcloud.

The libraries that are used for generating wordcloud are:

- library (wordcloud): It is a function to make attractive word clouds, see the dissimilarity and resemblance among documents, and stay away from over-plotting in scatter plots with manuscript.
- library (RColorBrewer): Provides colour schemes for maps and other graphics. Creates nice looking color palettes especially for thematic maps
- library (syuzhet): Extracts sentiment and sentiment-derived plot arcs from text using a variety of sentiment dictionaries conveniently packaged for consumption by R users. Implemented dictionaries include "syuzhet" (default) developed
- library (reshape2): Reshape2 is a reboot of the reshape package. It's been over five years since the first release of reshape, and in that time. Reshape2 utiizes that information to create a novel package for reshaping data that is a great deal additional focused and much faster.

Visualization

We can easily infer from the bar plot that the chat had a maximum number of positive sentiments followed by negative as second and anticipation at third.

Though Sentiment analysis has been the most popular textual analysis tools among businesses, scholars and analysts to take decisions and for research purposes Sentiment analysis has its own limitations as language is very complex and the meaning of each and every word changes with time and from person to person. Also, the accuracy of the analysis can't be accurately measured and compared with how human beings analyze emotions.

The libraries that are used for plotting a graph are:

- library (ggplot2): ggplot2 is a system for declaratively generating graphics, based on the grammar of graphics. If date is provided, ggplot2 maps variables to aesthetics, graphical primitives to utilize, and it takes concern of the details.
- library (scale): Scale is a practical website dedicated to clinical assessment scales.

```
> head(d, 10)
                      word freq
jagadeeshwari jagadeeshwari   40
pranudas           pranudas   32
the                     the   21
mam                     mam   20
you                     you   20
are                     are   13
media                 media   13
omitted             omitted   13
with                    with    8
meet                    meet    8
> tail(d, 10)
           word freq
whatever whatever    1
which       which    1
work         work    1
problem   problem    1
deleted   deleted    1
message   message    1
ill           ill    1
output     output    1
were         were    1
lunch       lunch    1
```

Fig.4.2. Frequency of words

Fig.4.3. Wordcloud

```
> head(Sentiment)
  anger anticipation disgust fear joy sadness surprise trust negative positive
1     0            1       0    0   0       0        0     1        1        0
2     0            1       0    0   1       0        1     1        0        1
3     0            0       0    0   1       0        0     1        0        1
4     0            0       0    0   0       0        0     0        0        0
5     0            1       0    0   0       0        0     0        0        0
6     0            0       0    1   0       0        0     0        0        0
> tail(Sentiment)
   anger anticipation disgust fear joy sadness surprise trust negative positive
89     0            1       0    0   1       0        1     1        0        1
90     0            0       0    0   0       0        0     1        0        0
91     0            0       0    0   0       0        0     0        0        0
92     0            0       0    0   0       0        0     0        0        0
93     0            0       0    0   0       0        0     0        0        0
94     0            0       0    0   0       0        0     0        0        0
```

International Journal of Engineering Technology and Management Sciences[IJETMS]

Website: ijetms.in Issue:5, Volume No.4, September-2020 DOI: DOI: 10.46647/ijetms.2020.v04i05.014

Fig.4.4. Sentiment words

```
> TotalSentiment
      sentiment count
1         anger     1
2  anticipation    13
3       disgust     1
4          fear     4
5           joy     8
6       sadness     5
7      surprise     6
8         trust    17
9      negative     6
10     positive    16
```

Fig.4.5. Count of sentiment words

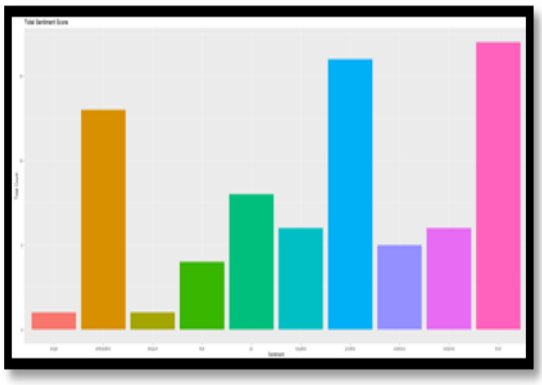

Fig.4.6. Total Sentiment Score

V TESTING

Software Testing:

Software testing is a critical element for software quality and assurance and represents ultimate review of specifications, design and implementation. Testing is an exposure of the system to trail input to see whether it produces correct output.

Testing Phases:

Software Testing includes the following:

1. Testing activities are determined and test data is selected.
2. The test is conducted and results are compared with the expected results.

Testing Activities:

Unit Testing:

This find faults by isolating an individual component using test stubs and drivers and by exercising the components using a test case.

We have performed unit testing on individual component and then integrated with the system that helped us to develop the efficient product.

Integration Testing:

This finds faults by integrating several components together. System testing, which focuses on the complete system, its functional and non – functional requirements and its target environment.

Whenever we integrate an individually working component with the system integration testing helps to find the compatibility with the system.

Equivalence Testing:

It is a black box testing technique that minimizes the number of test cases. The possible inputs are separated to correspondence testing is that the system usually acts in similar ways for all member of a class.

VI CONCLUSIONS & FUTURE SCOPE

Conclusions

From the performed analysis and visualization, it is found that the sentiments that are utilized in the WhatsApp chat. So as to conclude WhatsApp is the best communication platforms whose pros and cons are decided by the user itself. This has been shown by analyzing and visualization the logins and actions happened in the groups, assessing the information of their consideration through Word Cloud analysis and measuring the sentiments of the texts/words used within the chat. In this project, chat of WhatsApp has been used as database by using R, sentiments and emotions are being analyzed. Thus, it has been proved that number of techniques can be used to perform the sentiment analysis. Each technique has its specific domain. If used positively then it's a boon for the users and conditionally addicted after that a ban and therefore this paper, classified the stage of addiction of users to the chat of the WhatsApp so as to limit the time spend on it and to explore the chat whenever necessary.

Future Scope

Hence, future scope of sentiment classification domain is the accuracy and efficiency. The proposed methodology as of now giving good accuracy but in many texts Hindi words are also included which has no sentiment value so in future research can work on that and also will work on the data filtering to get the meaning of non-English and stress words.

REFERENCES

[1]. Seufert, M., HoBfeld, T., Schwind, A., Burger, V., Tran-Gia, P.: Group-based communication in WhatsApp. IFIP Networking (2016)

[2]. Thakkar, H., Patel, D.: Approaches for sentiment analysis on Twitter: a state-of-art-study. In: International Network for Social Network Analysis Conference (INSNA), Xi'an China, July 2013

[3]. Torgo, L.: Data Mining with R Learning with Case Studies. CRC Press, Taylor & Francis Groupan Informa Business (2011)

[4]. D'Andrea, A., Ferri, F., Grifoni, P., Guzzo, T.: Approaches, tools and applications for sentiment analysis implementation. Int. J. Comput. Appl. 125(3) (2015)

[5]. Fang, X., Zhan, J.: Sentiment analysis using product review data. J. Big Data (2015) (A SpringerOpen Journal)

[6]. Hussein, D.-M.E.D.M.: A survey on sentiment analysis challenges. J. King Saud Univ. Eng. Sci. (2016)

[7]. Joshi, K., Patel, D., Pandya, S.: A survey on sentiment analysis techniques. Int. J. Innov. Res.Technol. 3(7) (2016)

[8]. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. 5 (2014) Sentiment Analysis on WhatsApp Group Chat Using R.

Website: ijetms.in Issue:5, Volume No.4, September-2020 DOI: DOI: 10.46647/ijetms.2020.v04i05.014

[9]. Neri, F., Aliprandi, C., Capeci, F., Cuadros, M.: Sentiment analysis on social media. In: IEEE/ACM International Conference on Advances in Social Networks: Analysis and Mining (2012)
[10]. Pradhan, V.M., Vala J., Balani, P.: A survey on sentiment analysis algorithms for opinion mining. Int. J. Comput. Appl. 133(9) (2016)
[11]. Safrin, R., Sharmila, K.R., Subangi, T.S., Vimal, E.A.: Sentiment analysis on online product review. Int. Res. J. Eng. Technol. 04(04) (2017)
[12]. Shaikh, A., Rao, M.: Survey on sentiment analysis. In: International Conference on Emanations in Modern Technology and Engineering (ICEMTE-2017), vol. 5, issue 3 (2017)